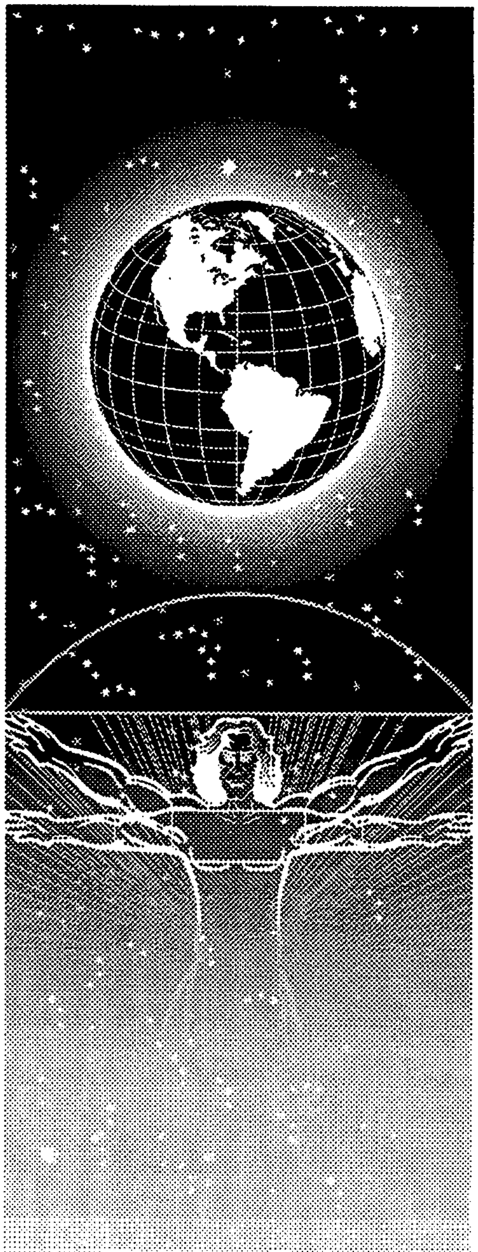
AL/CF-TR-1997-0133



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

SUBJECTIVE ASSESSMENT OF SAR IMAGERY ENHANCEMENT ALGORITHMS

Gilbert G. Kuperman

CREW SYSTEMS DIRECTORATE
HUMAN ENGINEERING DIVISION
WRIGHT-PATTERSON AFB OH 45433-7022

Dorit Shaya

HUMAN FACTORS ENGINEERING BRANCH
ISRAEL AIR FORCE

SEPTEMBER 1997

INTERIM REPORT FOR THE PERIOD JANUARY 1997 TO SEPTEMBER 1997

Approved for public release; distribution is unlimited.

Crew Systems Directorate
Human Engineering Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1997-0133

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

John F Kent

JOHN F. KENT, COL, USAF, BSC
Acting Chief,
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1997	Interim, Jan - Sep 97

4. TITLE AND SUBTITLE
Subjective Assessment of SAR Imagery Enhancement Algorithms

5. FUNDING NUMBERS
PE 62202F
PR 7184
TA 10
WU 44

6. AUTHOR(S)
Gilbert G. Kuperman
*Dorit Shaya

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
*Israel Air Force Human Factors Engineering Branch

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Crew Systems Directorate
Human Engineering Division
Human Systems Center
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/CF-TR-1997-0133

11. SUPPLEMENTARY NOTES
This effort was sponsored by AFMC/ST under the International Opportunities Fund Program.

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Image display enhancement algorithms employing Statistical processing and Fuzzy Logic processing were applied to 14 synthetic aperture radar (SAR) images. Three groups of subject matter experts, provided by the Israel Ministry of Defense (Air Force imagery analysts, Ground Corps Command imagery analysts, and Air Force weapon system officers) employed a subjective image interpretability rating scale in assessing the affect of the enhancement processing. No statistically significant effect was found for either the algorithms or the groups of subjects.

14. SUBJECT TERMS
Image Interpretability, SAR, imagery analyses, Human Factors, target acquisition

15. NUMBER OF PAGES
23

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

This page intentionally left blank.

PREFACE

This study was conducted under the auspices of Data Exchange Annex DEA-IS-8703, "Human Factors and Biodynamics," between the air forces of the United States of America and the State of Israel. The execution of the research was supported by an Air Force Materiel Command, Directorate of Science and Technology (AFMC/ST), International Opportunities Fund (IOF) award. The scope of the IOF-supported research includes the cooperative evaluation of both display enhancement algorithms and sensor fusion algorithms. The current study, evaluating display enhancement algorithms for application to synthetic aperture radar (SAR), was conducted with the joint participation of the Israel Air Force (IAF) Human Factors Engineering Branch (HFEB) and the Human Engineering Division of the United States Air Force (USAF) Research Laboratory, Wright-Patterson Air Force Base, Ohio. The research took place in Israel during the period 20 July 1997 through 28 August 1997, under the auspices of Data Exchange Annex DEA-8703 ("Human Factors and Biodynamics"). The USAF participation was under Work Unit 71841044, "Advanced Crew Systems for Reconnaissance, Surveillance and Target Acquisition."

The authors wish to thank Mr. Roger Cranos of the (United States) Air Force Research Laboratory (WL/AAZI), the Technical Project Officer for DEA-8701 (Avionics), who provided the SAR imagery, and to Mr. Michael Avraham (of Elta) who carried out the image enhancement processing which was evaluated during the experiment. Special thanks are due to Sgt Amir of the HFEB for his unflagging support in preparing and supporting the experimental apparatus.

In any study of this sort, the true knowledge resides with the subject matter experts of the operational units. The authors are indebted to the young men and women of the IAF Imagery Analysis Unit, the Ground Corps Command Imagery Analysis Unit, and the Israeli Air Force Fighter Squadrons for their participation in support of the study.

The USAF author expresses his sincere appreciation to the staff of the IAF HFEB. Their high motivation and professionalism made his work in their laboratory a truly memorable experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page intentionally left blank.

SECTION I

INTRODUCTION

Background

Many military reconnaissance and target acquisition tasks are performed under high levels of stress. An imagery analyst (IA), for example, may have discovered a theater mobile missile system or a fighter aircraft may be conducting a night, in-weather ground attack against a defended target. In the first case, the IA must report this target immediately because the missile might be capable of delivering chemical, biological or nuclear warheads. In the second example, the aircrew seeks to minimize the effectiveness of the enemy's defensive systems by making a single pass attack, but must also ensure against fratricide and seek to minimize any possible collateral damage in the vicinity of the assigned target. In both cases, the task might well rely on a standoff, all-weather imaging sensor such as synthetic aperture radar (SAR).

SAR imagery is very different from optical photography or electro-optical sensor imagery such as television. Intensities in the SAR imagery are related to the radar cross section of objects being imaged (as opposed to the visible wavelength spectral reflectance of those objects). Resolution (the ability to separate two closely spaced objects in the imaged scene, or the fineness of the detail which may be discerned from the image) of SAR imagery is generally much lower than that of these other sensors and may be on the order of meters (as opposed to cm or mm).

SAR, however, offers unique advantages compared to sensors whose imagery is more literal. It can be employed at great standoff ranges, out to 100s of km. It is independent of time of day and provides imagery of constant quality during day or night. It can also be employed under all but the most severe of weather conditions without suffering significant image quality degradation.

Because of this combination of reasons, the air forces of both Israel and the United States have exhibited a great deal of interest in employing SAR systems for a variety of missions while seeking to improve the utility of SAR sensor imagery. The USAF is exploring enhancements to the SAR modes of both the F-15E Strike Eagle and the E-8C Joint Surveillance and Target Acquisition Radar System while the IAF is considering improving the resolution of the multi-mode radar on their F-4 Phantom 2000 multirole fighter. Another (and less expensive) approach to achieving improved operational capabilities is to apply image processing algorithms to the imagery. This is expected to result in improvements in the speed and/or accuracy with which the operator can extract information from the imagery.

Elta Electronics Industries, Ltd., the electronics and avionics subsidiary of Israel Aircraft Industries, Ashdod, developed proprietary algorithms for the display enhancement of SAR reconnaissance, surveillance and target acquisition imagery. The algorithms were previously applied to detected digital imagery and were reported, during a pilot/operational utility demonstration activity, to show promise for improving the interpretability of the imagery for exploitation by a weapon system officer (WSO) in a target acquisition task or by an IA performing a reconnaissance/surveillance task. The algorithms were reported by Elta to both improve the imagery signal-to-clutter ratio by reducing speckle and to improve the overall contrast of SAR imagery.

Objective and Approach

The objective of this study is to provide the first quantitative assessment of the Elta SAR image display enhancement algorithms. A subjective image interpretability rating scale, developed in the United States and widely used for both sensor system specification and testing, the Radar National Image Interpretability Rating Scale (RNIIRS), was selected and agreed to by both Air Forces' human factors engineering laboratories to serve as the performance metric. Enhancement processing of the SAR imagery set was performed by Elta. Subject matter experts (SMEs) were provided by the IAF. The study was conducted in Israel, jointly between the IAF and the USAF, at the facilities of the Israel Air Force (IAF) Human Factors Engineering Branch (HFEB).

SECTION II

METHOD

Imagery

The imagery used in this study was acquired by a developmental SAR radar system. This imagery contains a variety of tactical targets (vehicles - tanks and APCs - and a simulated AAA site) against a background of natural clutter.

Imagery Preparation

The 15 original SAR images were of size 2048 by 512 pixels and recorded in tiff (tagged image file format) at 8 bit/pixel or 256 gray levels. Each image was displayed on a SGi workstation using the commercial image manipulation package ImageWORKS. Fourteen 512 by 512 pixel subimages were extracted from the overall scenes by cropping (again using ImageWORKS). The subimages were recorded to disk in tiff/gray scale format. Each subimage contained activity of military interest: either deployed tactical vehicles or the simulated air defense site. These 14 extracted subimages were used as the "original" image set in the experiment.

The original image set (512 by 512 pixel size) was provided to Elta on digital cassette. All image enhancement processing was performed by Elta on the extracted original images. Each of the two enhancement algorithms (referred to as "statistical" and "fuzzy logic") was applied to the SAR imagery. This resulted in four imagery sets: Original (unenhanced), "statistically" enhanced using two window sizes of 5 by 5 and 7 by 7 picture elements (pixels) (referred to as "Statistical 5" and "Statistical 7," below) and "fuzzy logic" enhanced, using a window size of 7 by 7 pixels. A total of 56 stimulus images (14 images by four versions) was produced. After processing had been performed, all images were converted into XWD format for use in the experiment. Of the 14 scenes, 11 contained vehicles, one contained only the AAA site, and two contained both types of target.

Measure of Performance

A subjective utility evaluation tool, the Radar National Image Interpretability Rating Scale (RNIIRS [FAS, 1992]), developed by the United States specifically for the evaluation of SAR imagery for reconnaissance (and target acquisition) applications, was employed to evaluate the original and enhanced imagery. RNIIRS is based on the capability of imagery, at each specified image quality category, to support the extraction

of militarily-relevant information by a trained and experienced IA. RNIIRS is a ten category (zero through nine) rating scale. It provides exemplar verbal descriptions of the level of detail of the information extraction that should be supported by SAR imagery at that category. There are two basic types of verbal description included in the RNIIRS: "distinguishing" between possibly similar objects and "declaring" what an object is. The declaration can be at either of two levels: detecting or identifying. Detection refers to establishing the presence of the object or activity called out in the descriptor. For example, RNIIRS category level 3 requires that the imagery support reporting the presence (if any) of "vehicle revetments at a ground forces facility." Identification, on the other hand, refers to providing a specific name (or other label) to the detected object. For example, an image of RNIIRS level 7 is capable of supporting the correct assignment of the nickname FLOGGER to a small fighter aircraft. In the RNIIRS, the category level 0 rating is assigned to imagery which is so poor that "interpretability of the imagery is precluded by obscuration, degradation, or very poor resolution."

Interpretability rating scales are routinely employed in USAF acquisition programs both to specify the level of performance required from the sensor system and to test and evaluate actual performance. The Advanced Tactical Airborne Reconnaissance System Program, for example, used such a scale for these purposes.

The RNIIRS was translated into Hebrew to allow use by the Subject Matter Experts (SMEs). The translation was validated through review by an IAF intelligence officer/senior imagery analyst.

Subject Matter Experts

Both IAs and WSOs participated as SMEs in support of the study. Five personnel, one officer and four enlisted, two male and three female, assigned to the IAF Imagery Analysis Unit participated. Their ages were between 20 and 24 years. All had received three to four months training in imagery analysis and had experience in softcopy exploitation. All had received on-the-job training (OJT) in SAR exploitation. All had employed other subjective rating scales. These scales were for assessing "image quality" and had five reporting levels: bad [sensor malfunction or severely degraded] poor, fair, good, and excellent. All the IAF SMEs had some experience in exploiting SAR imagery (from the TERMIN sensor system) and each had received on-the-job training in the characteristics of SAR imagery and in SAR exploitation techniques.

Additionally, five IAs from the Ground Corps Command's (GCC) Imagery Analysis Unit also served as SMEs. They were all enlisted personnel, four male and one female, between the ages of 19 and 20. They were all experienced in SAR exploitation although from film-recorded imagery only. Two reported using qualitative rating scales to report image quality. All had experience with personal computers but only two reported

experience in performing softcopy exploitation. All had received service school training in imagery interpretation and OJT in SAR exploitation techniques.

Six experienced WSOs, all male officers, from the Israeli Air Force Fighter Squadrons, served as SMEs. (Currently, all combat air crews in the IAF are male.) They were between 20 and 24 years of age. All were experienced (between three and 50 flight hours) in using the SAR mode of the AN-APG-76 multimode radar, referred to as "Mitzpe Ram," for a variety of air-to-ground mission functions: navigation, target acquisition, target search, and weapon delivery.

Equipment

A SGi INDY graphics workstation was used both to present the imagery to the SMEs and to record their RNIIRS responses. The workstation employed a Model GDM 20E21 color monitor.

VAPS™ (Virtual Applications) version 4.1 (Virtual Prototypes Inc., Montreal, Canada), a commercial software utility for the conceptual design, testing, and evaluation of human-computer interfaces, was used to develop the task interface. Once the interface had been prototyped, the VAPS™ C code generator utility was employed to produce machine-independent code.

The brightness and contrast controls of the monitor were adjusted by the experimenter to support the display of a variety of information types: imagery, alphanumerics and graphics. The SME was not allowed to adjust these controls during data collection.

Task

"Ground School"

Upon arrival, each SME was provided with an orientation to the methodology of the study. Care was taken to explain that the SME was evaluating the interpretability of a set of SAR images (and not, themselves, undergoing testing). Each SME completed a brief background and experience questionnaire regarding their training and experience in the exploitation of SAR imagery. The RNIIRS rating scale was provided and its use was explained. The SME was instructed to evaluate each image independently of any others and to use the RNIIRS in as consistent a manner as possible. The SME was then seated at the graphics workstation and its functions (image display and response recording) were explained and demonstrated. The presence of corner reflectors (used for radar calibration and ground truth registration purposes) in the SAR imagery was pointed out and explained to each SME. When the SME reported that s/he was ready, the data collection portion of the experiment was initiated (by the SME).

Task Procedure

The procedure followed by each SME is depicted in Figure 1.

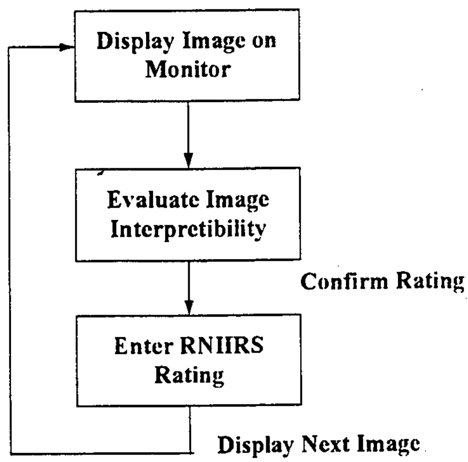


Figure 1. Flow Diagram of SME Task

Every image was displayed to each SME. The imagery was presented in a random order unique to each SME (to minimize order of presentation or learning effects). All imagery was displayed with the direction of radar illumination coming from the SME's left (or radar shadows falling to the right).

The SME was seated at the workstation. A SME identifier number was entered. A ten button keypad and the first image appeared on the screen. (The image appeared in a window of size approximately 15 cm by 15 cm, centered on the monitor with the keypad to its right.) A digital counter, located above the SAR image, informed the SME which of the 56 trials was currently in progress. The functions of the interface were explained. The SME was told that the radar illumination would always be from their left, with "shadows" always falling to their right.

The SME then decided on the RNIIRS category rating most appropriate to the currently displayed image (original or enhanced version) and entered that rating by using the mouse to drive the on-screen cursor over the selected keypad button and left mouse button to designate that button. (Thus, only integer responses were possible.) On designating a rating button, the button became highlighted, providing visual feedback as to the selection. Once a rating was entered, an "OK" button then also appeared. The SME could either change the rating (by designating another rating scale button) or enter the "OK" button (confirming the currently selected rating and terminating the present trial). No time limit was imposed on the rating process. Once the "OK" button was pressed, a

"NEXT" button appeared and a medium gray patch filled the image display window on the screen. The SME could command the next image at his own pace by activating the "NEXT" button. This process was repeated until all images had been evaluated. (The functioning of the SME interface for the rating task is shown in Table 1.)

Table 1. SME Interface for Interpretability Rating Task

Task Element	*Control Input*	*Display Response*	*Comment*
(E) = Experimenter *(S) = SME*			
(E) Enter SME ID No.	terminal keyboard	keypad, image, counter = 1	Begin experiment; generate random stimulus order; open data file
(S) Enter 0-9 rating	mouse	button highlighted; OK button appears	Collect data
-change rating	mouse	button highlighted	Modify response
-confirm rating	mouse & OK button	image goes to gray field; OK button disappears; NEXT button appears	Complete trial
(S) Initiate next stimulus	mouse & NEXT button	new image appears; NEXT button disappears; increment trial counter	Self paced
(S) Enter rating for last stimulus	mouse, keypad, OK button	Exit to UNIX	End experiment; close data file

At the conclusion of the 56 trials, the SME was asked if he had "noticed anything about the imagery?" This open-ended question was used to elicit SME comments regarding the enhancement processing treatment. The SME was then shown several examples of the four versions of one or two scenes and asked which version was preferred. A second question, dealing with the utility and ease of use of the rating scale instrument itself was also asked of each SME.

SECTION III

ANALYSIS

Analysis of Variance

A two-factor (imagery treatment [T] and SME Group [G]), Analysis of Variance (ANOVA) for a repeated measures experimental design was used to determine if there was a statistically significant difference between the four levels of enhancement processing or between the three groups of SMEs. Table 2 presents the ANOVA summary:

Table 2. ANOVA Summary

Source:	*df*	*MS*	*F*	*Significance*
Group	2	15.20	3.18	$p > .08$
Error	13	4.78		
Treatment	3	.03	.46	$p > .71$
Treatment x Group	6	.04	.61	$p > .72$
Error	39	.07		

As may be seen from this Table, neither the main effects of Group and Treatment nor the Group x Treatment interaction were statistically significant. The mean RNIIRS ratings for the four imagery treatment levels are portrayed in Figure 2 (error bars represent standard deviations around the mean). The three processed versions of the imagery set did not yield ratings which were different from those assigned to the original versions (or from one other).

Mean interpretability ratings for the IAF and GCC IAs as well as the WSOs are depcited in Figure 3. Although the GCC IAs tended to assign higher ratings to the imagery on average, this difference was not statistically significant.

The Group x Treatment interaction is plotted in Figure 4. As can be seen in the figure, the GCC IAs rated the Statistical 7 imagery higher than the Statistical 5 imagery, while the other two groups of SMEs rated it lower. This resulted in the reversed slope of the data graph for the GCC IA ratings compared to those of the other groups. As revealed by the ANOVA, however, this difference was not statistically significant.

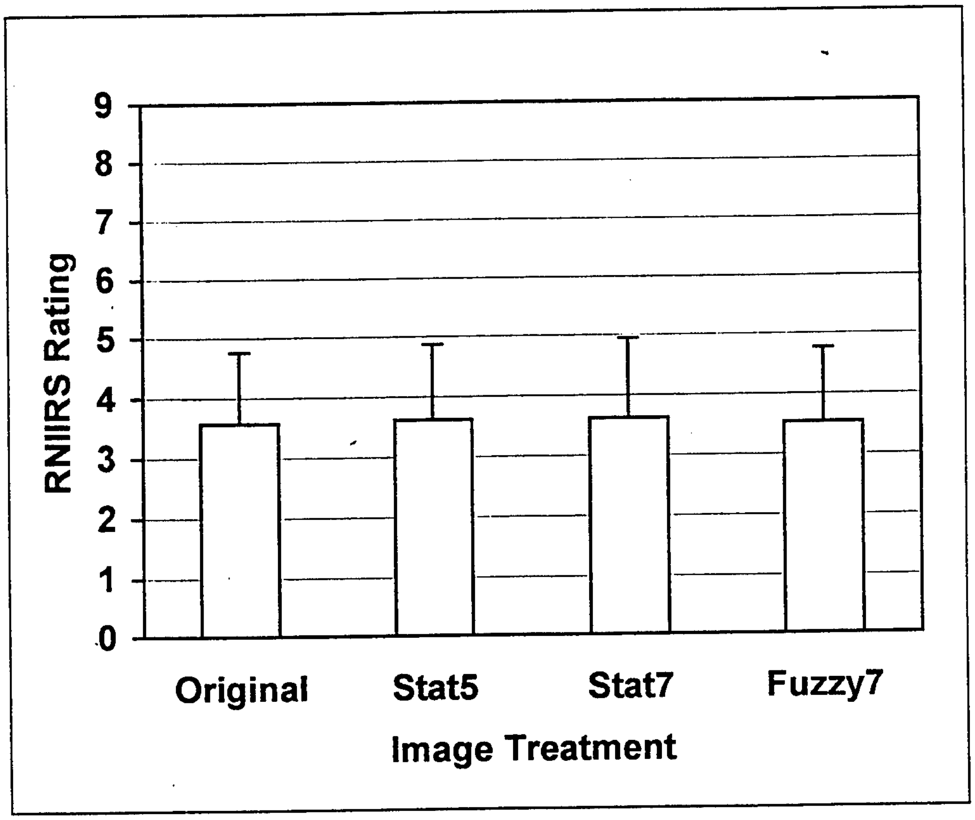


Figure 2. Mean Interpretability Ratings (and Standard Deviations) as a Function of Image Treatment

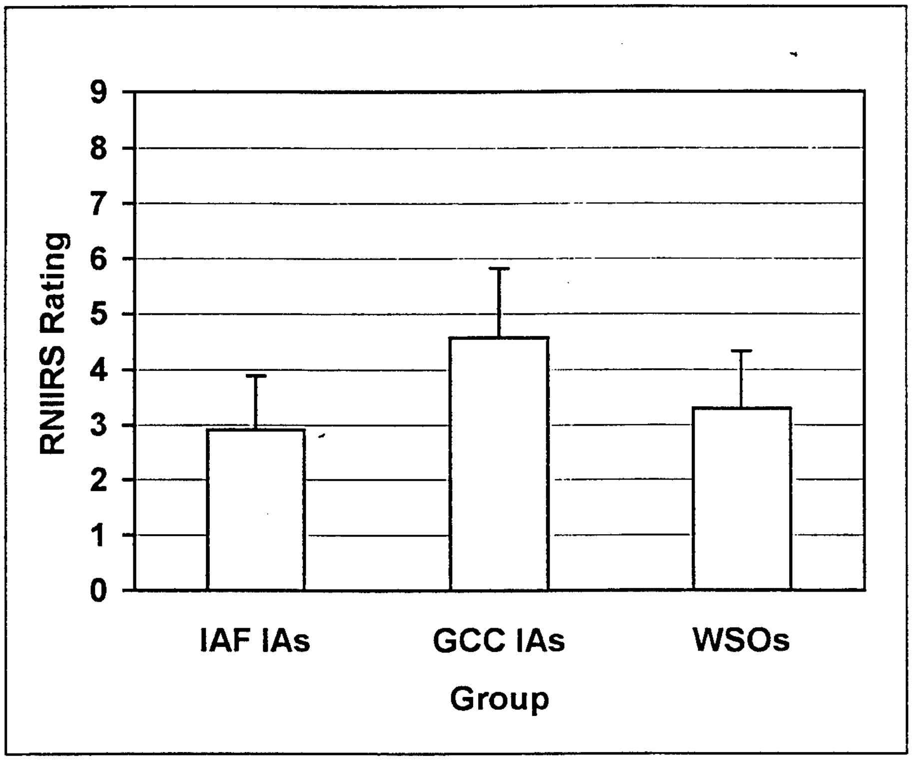


Figure 3. Mean Interpretability Ratings (and Standard Deviations) as a Function of SME Group

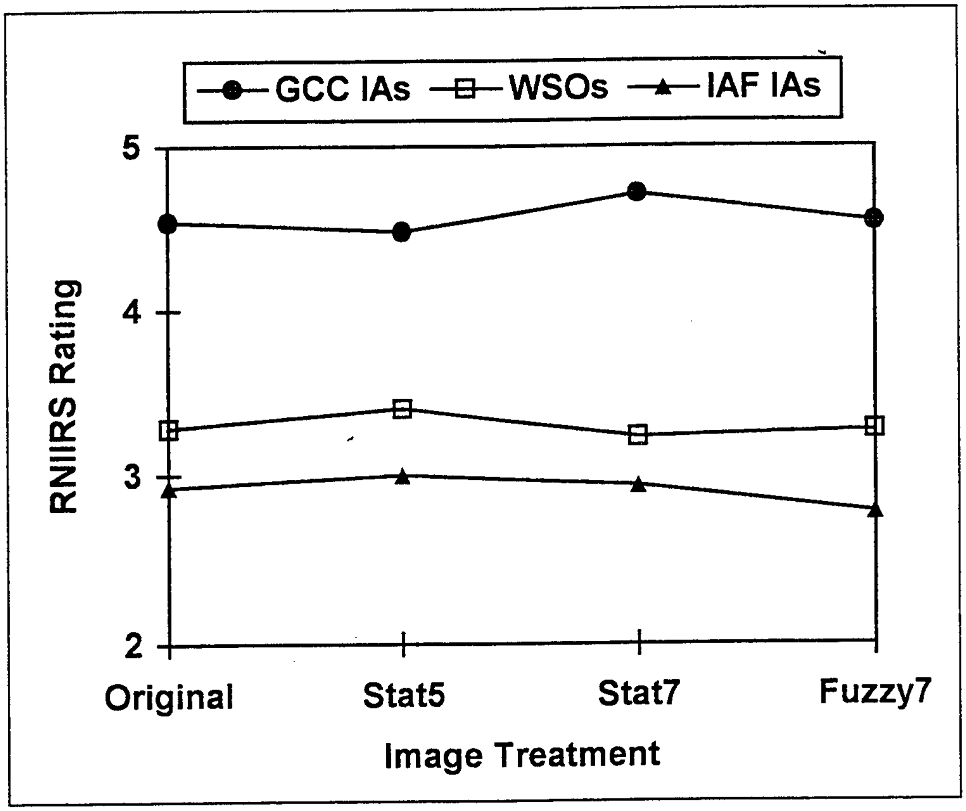


Figure 4. Mean Interpretability Ratings for the Image Treatment by SME Group Interaction (Expanded Scale)

As an additional verification of the experimental procedure, a check on order (or learning) effects was conducted by averaging the responses across all ten IA SMEs for the first four and last four stimuli. The means were: 3.725 and 3.750, respectively. This difference of 0.025 in the means has neither statistical nor operational significance.

SECTION IV

DISCUSSION

RNIIRS Ratings

No significant difference was found between the Imagery Treatments. The maximum difference in mean RNIIRS ratings was only 0.10 category levels. This difference was between the Statistical 5 processed imagery (mean RNIIRS = 3.62) and the Fuzzy 7 imagery set (mean RNIIRS = 3.52). This difference is too small to be considered to have operational meaning.

No statistically significant difference was found between the three Groups. It may be observed, however, that the GCC IAs assigned higher ratings (mean RNIIRS = 4.57) than did the other two Groups (mean RNIIRS for the IAF IAs = 2.91 and for the WSOs = 3.30). The difference is greater than a full rating scale step. This may reflect their (the GCC IA's) greater familiarity with exploiting imagery of ground order of battle targets.

SME Debriefing

IAs:

None of the 10 IAs (IAF and GCC, combined) reported any difficulty in using the rating scale. All noticed repetition of scenes within the stimulus set and, also, differences in sharpness and/or contrast (as illustrated in Table 3). Six preferred the Statistical 5 processed version, three preferred the Original version, and one preferred the Fuzzy Logic algorithm. All of the IAs explained their preference in terms of superior image sharpness, better definition of target boundaries and shadows, and/or enhanced contrast between the targets and their surround.

WSOs:

None of the WSOs reported having any difficulty in employing the RNIIRS as a rating instrument. Four of the six WSOs pointed out what they believed to be inconsistencies in the descriptors associated with several of the rating levels. One of them stated that the rating scale was irrelevant to his tasks since only the contrast and sharpness of the imagery affect target acquisition. Five WSOs reported that they had noticed both a repetition of the target scenes and differences between the images. They described the difference in terms of some images appearing to exhibit greater "clarity," of being better "focused," or having more "sharpness." Two of the WSOs expressed preference for the

Statistical 5, two for the Statistical 7 filtered imagery, one for the Fuzzy Logic filtered version, and one for the Original version.

Table 3 presents the results of the post test preference selection:

Table 3. SME Preferences by Imagery Treatment

Imagery Treatment:	Original	Statistical 5	Statistical 7	Fuzzy Logic
IAF IAs	1	3	0	1
GCC IAs	2	3	0	0
IAF WSOs	1	2	2	1

SECTION V

CONCLUSIONS AND RECOMMENDATIONS

The specific enhancement algorithms used in this study failed to demonstrate an improvement in either IA or WSO performance, at least as predicted by the RNIIRS instrument. (This may be due to a variety of experimental factors including the imagery employed, the types of targets imaged, *etc.*) This does not suggest that other algorithmic approaches may not result in dramatic improvements in IA and/or WSO performance. Research should be continued to develop and evaluate enhancement algorithms.

Three quarters of the SMEs expressed a preference for one or another of the enhanced imagery sets (over the Original) with half of all SMEs preferring the imagery processed by the Statistical 5 algorithm. Although not supported by improved image interpretability ratings, this preference may suggest that the SMEs were more confident when evaluating (and presumably exploiting) this imagery because they felt that they were better able to extract targets and target details. This preference may have operational significance when information must be extracted from SAR imagery under time constraints.

The check for a significant order or learning effect validated the experimental procedure. Careful explanation of the rating scale instrument, together with instructions which stressed its consistent application during the task, and SME self-report of understanding the use of the rating scale, were sufficient to avoid this potential confound.

REFERENCES

Federation of American Scientists (FAS), 1992, http://www.fas.org/irp/imint/niirs

GLOSSARY

AAA	Antiaircraft artillery
AFMC	(United States) Air Force Materiel Command
DEA	Data Exchange Annex
GCC	Ground Corps Command
HFEB	Human Factors Engineering Branch (of the IAF)
IA	Imagery analyst
IAF	Israeli Air Force
IOF	International Opportunities Fund
OJT	on-the-job training
pixel	picture element
RNIIRS	Radar national imagery interpretability rating scale
SAR	Synthetic aperture radar
SGi	Silicon Graphics (workstation)
SME	Subject matter expert
tiff	tagged image file format
USAF	United States Air Force
WSO	Weapon system officer